二分类文档

1. 实现思路

使用浅层神经网络进行训练，因为任务比较简单，如果网络结构过于复杂及其容易导致过拟合，使用投票的方式进行模型的测试，因为一个音频不仅仅只有一种类型，因此需要使用投票的形式。

1. 网络结构

简单的全连接，使用pytorch框架和39维\*5帧的mfcc特征作为输入进行分类任务。后期的思路是改进为一维卷积的形式进行训练，提取高维语义特征使得可分。

1. 目前问题

准确率，和泛化能力较差。

1. 改进策略

加大数据量和增加特征。

数据输入：

数据集的制造，特征的提取。

特征的大小。

64, 39, 20 ： 64个sample，每个smaple 20帧，39特征MFCC。

使用 Torch.utils.DataLoader加载数据，使用Dataset制作数据集。

网络层：

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| input | torch.Size([64, 39, 20]) | 64 |  |
| C1 | 39\*1024 | torch.Size([64, 1024, 19]) | 卷积 |
| C2 | 1024\*512 | torch.Size([64, 512, 18]) |
| C3 | 512\*256 | torch.Size([64, 256, 17]) |
| C4 | 256\*128 | torch.Size([64, 128, 16]) |
| C5 | 128\*64 | torch.Size([64, 64, 15]) |
| F6 | 64\*15, 128\*2  128\*2 | torch.Size([64, 256]) |  |
| F7 |  | torch.Size([64, 128]) |  |
| F8 |  | torch.Size([64, 2]) |  |

优化器：

测试：

语音增强：

数据集的制造：